音声研究　第14巻第１号
2010(平成22)年４月
35–47頁

Journal of the Phonetic Society
of Japan, Vol.14 No.1
April 2010, pp.35–47

The Influence of Lexical Factors on Word Recognition by Native English Speakers and Japanese Speakers Acquiring English: A First Report

Kiyoko YONEYAMA* and Benjamin MUNSON**

英語母語話者と日本人英語学習者の語認識に与える語彙要素の影響についての報告

米山　聖子＊，ベンジャミン・マンソン＊＊

要旨：本論文は英語を学習している日本人成人の音声語彙認識における語彙頻度，語彙親密度そして音韻的近傍密度の影響の研究に関する最初の報告である。研究には，英語力が低い東京または埼玉在住者，英語力が高い米国ミネアポリス在住者，そして米国ミネアポリスに在住の英語母国語話者の３つのグループの聴取者が研究に参加した。Imai, Walley, and Flege（2005）に従い，英語母国語話者もしくは日本語母国語話者で日本語アクセントのある英語を話す話者が発した音韻的近傍密度と語彙頻度を直行に変化させた80語を聴取者に提示した。音韻的近傍密度と話者の母語言語の強い影響がすべてのグループで見られた。その上，英語力が高い日本人聴取者は他の２つのグループよりも，アクセントを持つ話者が発した語を聴取する場合の知覚の減少が少なかった。しかしながら，刺激語を厳密に分析してみると，母音と子音は条件にランダムに分布しているわけではなく，非英語母国語話者が犯した特定のエラーは近傍密度の高い語に偏って現われている母音に起こることが明らかになった。今後の研究への提案も行う。

Key words: spoken-word recognition, word frequency, word familiarity, phonological neighborhood density

1. Introduction

A central endeavor in the field of spoken language is the understanding the factors that predict how well individuals learning a second language (L2) can perceive and comprehend that language. A great deal of research in L2 acquisition has examined this question at the level of phoneme perception. A very consistent finding in this literature is that L2 phoneme perception is affected principally by two factors. The first of these is the age at which the L2 is acquired, with early acquisition favoring native-like phoneme perception (Flege and Liu 2001, Yamada 1995).

The second is that phoneme perception is affected by the degree of mismatch between the phonemic system of the first and second language. L2 phonemes that can be assimilated to a single phonemic category in the L2 listener's L1 present the greatest discrimination difficulty; consequently, adult L2 learners are often not sensitive to all of the relevant contrasts in the L2. One straightforward case is the difference between the Spanish and English vowel systems. Spanish has a five-vowel system, /i/, /e/, /ɑ/, /o/, /u/, while the system of English is much larger. Spanish listeners generally perceive English /i/ and /ɪ/ as instances of Spanish /i/ (Best 1995). This is not surprising, given that the English phonemes' formant frequencies cover the same space for the single /i/ category in Spanish. Numerous other examples of this can be found: English listeners perceive Mandarin /ɕ/; and /ʂ/ as instances of English /ʃ/ (Li 2008). Similarly, word-initial /p/, /t/, and /k/ are produced with more aspiration in English than Italian, and word-final stops are less likely to be released in English than in Italian. Italian-speaking L2 speakers of English have difficulty differentiating among places of articulation when listening to English word-initial and word-final /p/, /t/, and /k/ (MacKay, Meador, and Flege 2001). Even the perception of bilinguals whose language performance is relatively equivalent in normal tasks perceived phonemes differently in more-challenging tasks. In those tasks, phoneme perception is affected by a dominant-language phonemic system. A gating experiment (in which words are presented with acoustic information removed) has shown that

* Associate professor, Department of English, Faculty of Foreign Languages, Daito Bunka University（大東文化大学外国語学部英語学科准教授）
** Associate professor, Department of Speech-Language-Hearing Sciences, University of Minnesota（ミネソタ大学音声言語聴覚科学科准教授）

Spanish-dominant 'balanced' Spanish-Catalan bilinguals require more phonetic information to identify phonemes that exist in Catalan but not in Spanish than Catalan-dominant bilinguals do. (Sebastian-Galles and Soto-Faraco 1999).

The focus of the current investigation is on perception of spoken words, rather than phonemes *per se*. This investigation is motivated by previous studies with English L1 listeners that have found that lexical properties such as word frequency, word familiarity and phonological neighborhood density affect spoken word recognition. More frequent words and more familiar words are recognized quickly and accurately than less frequent and familiar words. Phonological neighborhood density is a measure of the similarity of the phonological words stored in the lexicon (e.g., Luce and Pisoni 1998, Vitevitch and Luce 1999). Phonological neighborhood density is conventionally calculated by counting the number of phonologically similar words within the range of one-phoneme edit distance (i.e., real words that can be made by inserting, substituting or deletion a single phoneme from a target word). Therefore, example neighbors for the word *cat* are the words, *scat*, *sat* and *at*, among others. There are created by phonological operations of insertion, substitution and deletion, respectively. Previous studies have shown that words with few phonological neighbors (i.e., those from a sparse neighborhood) are recognized more quickly and accurately than words with many neighbors (i.e., those from a dense neighborhood). Research on L1 acquisition has shown that neighborhood density effects on language behavior change over development. Storkel (2002) showed that children's representations for words in dense neighborhoods were more segmental than those in sparse neighborhoods. Munson, Swenson, and Manthei (2005) showed that the effect of phonological neighborhood density on response latencies in a real-word repetition task was not present in younger children (whose average age was 4 years) but was present in older children (whose average age was 7 years). These findings and others are consistent with Metsala and Walley's (1998) claim that vocabulary growth changes the nature of phonological representations, such that the initially relatively holistic representations become more segmentally structured.

Bradlow and Pisoni (1999) investigated whether word frequency and phonological neighborhood density affect word recognition in L2 listeners who spoke a variety of first languages. They presented English words in noise to native and non-native listeners and asked them to write what they heard. The word accuracy data revealed that there was a bigger discrepancy in recognition accuracy between easy English words (i.e., those with high frequency of use and low neighborhood density) and difficult English words (those with low frequency and high neighborhood density) by L2 listeners when compared with L1 listeners, such that the L2 listeners were disproportionately poorer than L1 listeners on 'hard' words. This suggests that suggesting that non-native listeners experience a particular challenge in using fine phonetic information to identify sounds in challenging-to-perceive words.

Further evidence for a relationship between phoneme perception and spoken-word recognition is presented by Meador, Flege, and MacKay (2000). These investigators presented low-predictability English sentences in the presence of background noise to early and late Italian-English bilinguals and measured repetition accuracy. The results showed that perception of English vowels consonants predicted a significant proportion of variance in word recognition, beyond what was accounted for by other factors known to affect word-recognition accuracy, such as length of exposure to English. This finding suggests that phoneme perception in L2 also plays a key role in L2 spoken word recognition. This study shows a close relationship between L2 learners' segmental perception and word recognition.

Imai, Walley and Flege (2005) conducted a series of experiments to reveal the effects of lexical properties in L2 spoken-word recognition. Imai et al.'s study was an extension of Bradlow and Pisoni (1999). As mentioned above, Bradlow and Pisoni tested native and non-native listeners, but the language background of the non-native listeners was not well controlled. Listeners' first languages were typologically diverse, including Korean, Mandarin, Russian, Japanese, and Spanish. Given the great variation in these languages' phonological and lexical characteristics, there may have been critical patterns of L1 interference that were missed.

Imai and colleagues investigated Spanish L1 listeners' word recognition in English, their second language. Specifically, they tested high- and low-proficiency L1 Spanish/L2 English listeners' perception of easy and hard English words in noise. The stimulus words varied orthogonally in their word frequency and phonological neighborhood density. The stimuli were produced either with Spanish-accented English produced by a native Spanish speaker or with natural English produced by a native English speaker. The three groups of listeners participated, those of whom were high-proficiency L1 Spanish/L2 English listeners, those of who were low-proficiency L1 Spanish/L2 English listeners, those of whom were native English listeners. Both Spanish-accented and natural English words were presented to the three groups of listeners and they were asked to write down what they heard. Spanish-accented stimuli were included under the assumption that less-proficient L1 listeners would recognize these productions more accurately than native-accented ones, as they match more closely the productions that the talkers make themselves.

The results showed that low-proficiency L1 Spanish/L2 English listeners were equally accurate in their recognition of Spanish-accented and naturally-produced English words. High-proficiency L1 Spanish/L2 English listeners and native listeners were less accurate for the Spanish-accented words; however, both groups of the L1 Spanish/L2 English listeners showed a smaller discrepancy than the native English listeners. All three groups were less accurate in recognizing difficult words (those with low frequency of use and high neighborhood density) than easy ones (those with high frequency of use and low neighborhood density). A reanalysis of the data revealed that word familiarity was a better predictor of performance differences across words than was word frequency.

The Imai et al.'s study provided three interesting findings. The first finding is that this study confirmed the claim by Bradlow and Pisoni: lexical properties (word frequency and phonological neighborhood density) affect word recognition not only in a native language but also in a second language.

The second finding is that word familiarity served as a better lexical factor than word frequency. The word frequency information Imai et al. used was based on texts. Recent studies have shown that word familiarity affects more on word recognition than word familiarity (e.g., Sekiguchi 2006 and among others). This study also further provided a supporting evidence for the effect of word familiarity.

The last finding is that L2 proficiency is highly related to the development of L2 lexical representations. The results seem to suggest that English language proficiency of Spanish listeners is a key to determine whether a Spanish-accented English representation or a more native-like English representation is used by non-native speakers. In L1, over learning, holistic representations change to more detailed segmentally better-specified representations (Charles-Luce and Luce 1990). We would expect that a similar transition would be developed in L2. With greater L2 exposure and additional word learning, L2 learners' lexical representations may also become more fine-grained or fully specified *vis-a-vis* their new language. As Flege, Munro, and MacKay (1995) have reported, good L2 pronunciation corresponds to more native-like L2 lexical representation, English lexical representations of high-proficient L1 Spanish/L2 English listeners were flexible enough to accommodate both Spanish-accented and native-like pronunciations.

The Imai et al.'s study is one of a few studies who have investigated how lexical properties of words stored in the lexicon affect non-native word recognition processes. In the conjunction to the results of previous L2 phoneme perception studies, L2 word recognition is affected by both sublexical and lexical information. In order words, phoneme perception at the prelexical level and lexical properties of words such as phonological neighborhood, word frequency and word familiarity at the lexical level are all important in the course of word recognition in L2. However, there still remain unsolved research questions on L2 word recognition.

The purpose of this paper is twofold. The first research question is related to the lexical-level processes of L2 word recognition. We will investigate whether Imai et al. (2005)'s findings can be extended to a second case of L2 acquisition, that of Japanese L1 speakers' acquisition of English as an L2. More specifically, the effects of neighborhood density, word frequency, and/or word familiarity will be tested. The effects of lexical properties such as neighborhood density and word frequency are attested not only in English, but also in other language such as in Japanese (Yoneyama 2002). It is also notable that there is evidence that the influence of neighborhood density on word-recognition and word production in the L1 studied by Imai et al., Spanish, is different from that in English (Vitevitch and Rodriguez 2005, Vitevitch and Stamer 2006). Therefore, it is also possible for L1 Japanese/L2 English listeners to employ lexical information while they listen to English.

The second research question is related to the prelexical level processes of L2 word recognition. We will investigate whether apparent effects of phonological neighborhood density can be attributed to the phonetic content of the stimuli rather than to lexical characteristics. The lists of words in Imai et al. were not balanced phonetically. As shown in the previous L2 perception studies, accurate phoneme perception plays an important role in L2 word recognition. Therefore, there might be a possibility that apparent effects of lexical difficulty can be attributed to difficulties perceiving specific phonemes.

In order to answer two questions, we try to replicate

Imai et al. Though we will argue that Imai et al.'s methodology is not perfectly suited to the specific questions that we have about Japanese L2 learners of English given the non-random distribution of sounds across different levels of frequency and neighborhood density, we feel that it's important to investigate whether we can replicate their findings using a typologically different L1. Therefore, we try to conduct our experiment as similar as possible to the one in Imai et al. (2005).

2. Methods

2.1 Subjects

Sixty-nine individuals participated in this study. All participants were aged between 19 and 50 with no hearing difficulties. There were three groups of participants: Minnesota English, Minnesota Japanese, and Tokyo Japanese. The Minnesota English participants were 22 speakers of English living in twin cities, MN and were undergraduate students, graduate students, and staff at University of Minnesota. The Minnesota Japanese participants were 24 native speakers of Japanese with high proficiency in English living in twin cities, MN and were mainly graduate students, postdoctoral fellows, staff and faculty members at University of Minnesota. The Tokyo Japanese participants were 23 native speakers of Japanese living in Tokyo or Saitama and were undergraduates at Daito Bunka University who had no experience living in the English speaking countries more than three months. Their English proficiency was low-intermediate (all but one students, TOEFL <475 at the time of study participation).

2.2 Stimuli

Stimuli in the word-recognition task were productions of the same word list that Imai et al. used. The summary of the stimuli is given in Table 1. For the complete word list, see Imai et al. (2005). The stimulus words were orthogonally manipulated with word frequency and phonological neighborhood density and fall into four word categories: high-frequency high-density words (HFHD), high-frequency low-density words (HFLD), low-frequency high-density words (LFHD), and low-frequency low-density words (LFLD). It is noteworthy to point out that the distribution of phonemes was not random across different frequencies and densities. Chi-squared contingency tests showed the distribution of onset consonants to differ marginally between high- and low-density words ($\chi^2_{[df=20]}$=30.367, p=0.064). Vowel nuclei differed significantly between high- and low-density words ($\chi^2_{[df=20]}$=28.262, p=0.029). The distribution of codas differed significantly as a function of both density ($\chi^2_{[df=20]}$=34.785, p=0.021) and, marginally word frequency ($\chi^2_{[df=20]}$=29.571, p=0.077). We return to this topic in the analysis section.

These words were spoken by two talkers. The first was a native speaker of English who was born in Michigan, had lived in Minnesota most of her adult life, and whose speech exemplifies some of the regional features unique to the North-Central US dialect region, particularly a very back-rounded /u/ and /oʊ/. The other was a native speaker of Japanese who teaches English in secondary school in Fukushima. Words were recorded with a high-quality microphone and were digitized at a 44.1kHz sampling rate with 16-bit quantization. Previous research has shown that word frequency and neighborhood density affect word duration (Munson and Solomon 2004, Gahl 2008, Yao 2010), we normalized the duration of the stimuli so that they were all 740ms long. This was done using the PSOLA algorithm in Praat (Boersma 2001). These were mixed with broadband noise at a +10dB signal-to-noise ratio. Stimuli for the word-familiarity task were printed versions of the 80 words used in the word-recognition experiment, along with 10 nonword fillers.

Table 1 A summary of the stimulus words

	HFHD	HFLD	LFHD	LFLD
Example	"bed"	"bring"	"bell"	"boss"
WF	195.5	177.8	18	22.4
ND	23.5	10	23.8	10.4

Note: WF: average word frequency; ND: average neighborhood density; HFHD: high-frequency high-density words; HFLD: high-frequency low-density words; LFHD: low-frequency high-density words; LFLD: low-frequency low-density words

2.3 Procedures

The word-recognition task was administered using the E-Prime experiment management software. Stimuli were played at a level of approximately 65dB SPL, through high-quality headphones. On each trial, a single word was played. There was a 100ms lag between the onset of the noise and the onset of the stimulus to allow listeners to become acclimated to the noise. The listeners then spoke their responses, immediately after which they typed their response in the response box. The purpose of eliciting these two responses was so that we could measure both response time (from the spoken response) and response accuracy (from the typed response). We did not want to use a spoken response alone out of concern that some of the Tokyo Japanese speakers' speech would be so heavily accented that it would be difficult to discern their target productions. Each listener heard each word only once. For each frequency/density combination, half of the words were produced by the native Japanese speaker, and half were produced by the native English speaker. Across the experiment, approximately equal numbers of participants heard each word spoken by the two talkers. The order of the stimuli was fully randomized, and each listener received a different unique randomization.

The familiarity task was similar to that in Imai et al., with ratings provided on a seven-point scale. Ours differed only in that the scale was inverted, such that 7 indicated that the word was heard often, and 1 that it was rarely heard. Following the familiarity task, the participants completed a sentence-production task, the results of which are not analyzed in this paper.

2.4 Analysis

The first analysis examines word-recognition accuracy. These were based on the written responses that participants gave. However, a subset of each participant's productions were transcribed orthographically and compared to the written responses. There was perfect agreement between these two; hence, written responses are analyzed here. Permissible misspellings of words (i.e., *cheeze* for *cheese*, *kart* for *cart*) were counted as acceptable responses. These were converted to a phonemic transcription, using the Carnegie-Mellon Pronouncing Dictionary. Two scores were calculated for each word. The first was a simple binary judgment of whether the person correctly reported the word. The percentage of words that people correctly recognized was calculated separately for each frequency/density combination produced by the native Japanese and native English speaker.

The second score on the percentage of phonemes in the target word that the person correctly reported. For this analysis, vowel-/r/ combinations were treated as a single diphthongal vowel. Unlike the analysis of percentage of words correct, participants could achieve 100% accuracy if they reported all of the phonemes in the word as well as additional sounds (i.e., reporting *cart* for target *car* would earn 100% in this analysis, but would be scored as 0 in the word-recognition score).

Individuals' average familiarity ratings for the written words were tallied separately for each frequency/density combination, and for the nonword foils.

3. Results

In this section, the results of three different analyses will be reported. The first part will deal with the results of the word-based analyses, followed by the part dealing with the results of the phoneme-based analyses. The final part will discuss the results of the analyses on our participants' word familiarity on the stimulus words.

3.1 Word recognition accuracy

The first analysis examined the relationship between lexical variables (word frequency and phonological neighborhood) and word-recognition accuracy. As in Imai et al. (2005), listeners' percent phonemes correct was submitted to a four-factor mixed-model ANOVA, with talker, frequency and density as the within-subjects factors, and group as the between-subjects factor. Significant effects of density ($F[1,66]=60.25$, $p<0.001$, $\eta^2_{partial}=0.47$), talker ($F[1,66]=6.73$, $p<0.001$, $\eta^2_{partial}=0.79$), and group ($F[2,66]=93.34$, $p<0.001$, $\eta^2_{partial}=0.73$) were observed. In addition, there were significant two-way interactions between frequency and group ($F[1,66]=3.77$, $p=0.028$, $\eta^2_{partial}=0.10$), between talker and group ($F[2,66]=32.45$, $p<0.001$, $\eta^2_{partial}=0.49$), between density and frequency ($F[1,66]=22.50$, $p<0.001$, $\eta^2_{partial}=0.254$), and between density and talker ($F[1,66]=8.05$, $p=0.006$, $\eta^2_{partial}=0.10$). Further, there was also a three-way interaction between frequency, density and talker ($F[1,66]=9.69$, $p=0.003$, $\eta^2_{partial}=0.12$). As shown in Fig. 1 and 2, for all participants, low density words were recognized more accurately than high density words. Words spoken by the English native speaker were recognized more accurately than the ones spoken by the Japanese native speaker. The Minneapolis English and the Minnesota Japanese equally recognized words accurately, but recognized words significantly more accurately than the Tokyo Japanese did. A large effect of frequency and a smaller effect of talker were observed for the Tokyo Japanese. A smaller effect of density was observed for low frequency words. The low frequency high density words were recognized less accurately when spoken by the native Japanese speaker than we would predict given the accuracy of low frequency high density words.

The words correct analysis clearly indicates that the word accuracy patterns were highly affected by the speakers who recorded the stimuli. Words for which the English native speaker was more than 10% less intelligible (averaged over all three groups of listeners) than the Japanese speaker were *soup, noon, date, face, lake,*

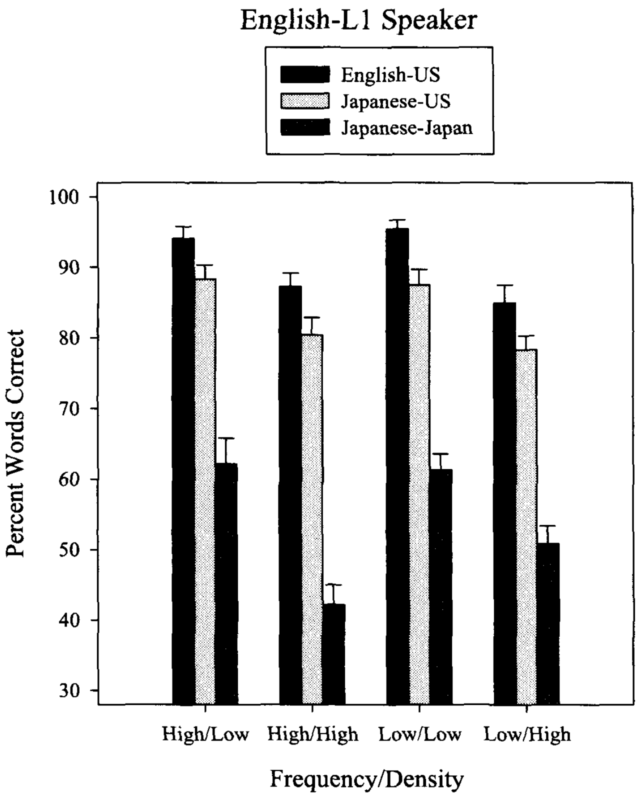

Fig. 1 Percent of words produced by the native English speaker that were correctly identified, separated by listener group.

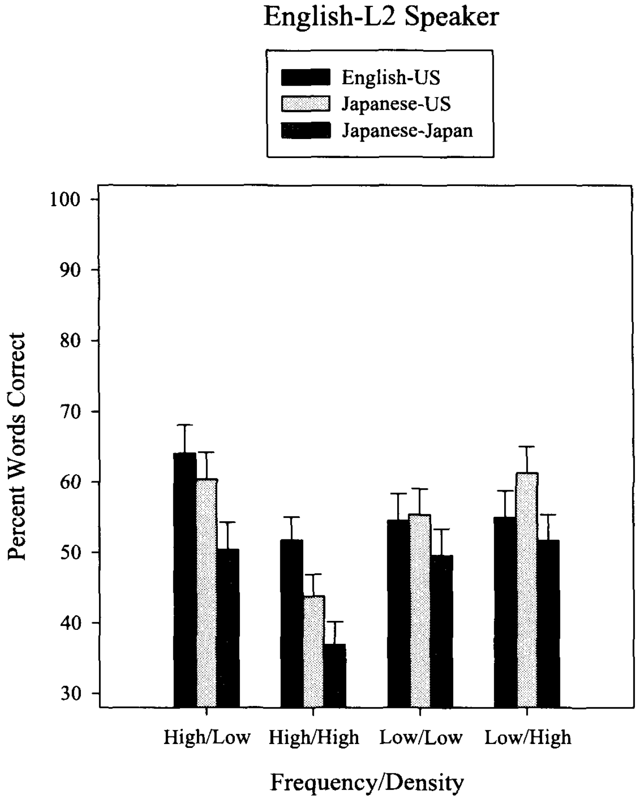

Fig. 2 Percent of words produced by the native Japanese speaker that were correctly identified, separated by listener group.

save, job, list, duck. Words for which the English native speaker was more than 50% more intelligible (average over all three groups of listeners) than the Japanese native speaker were *smell, teach, cute, match, mouth, note, corn, cart, part, cheese, boss, fall, fast, five, lost, park*. We further calculated average word-recognition accuracy of individual words produced by the English native speaker and the Japanese native speaker by the Minnesota English and the Tokyo Japanese listeners. We would expect that word accuracy data can be explained by lexical factors, such as the neighborhood density and word frequency. Words from sparse neighborhoods would be recognized more accurately than words from dense neighborhoods. However, there is no systematic relationship between the lexical variables and word-recognition accuracy for either of the talkers.

3.2 Phoneme recognition accuracy

In the word-accuracy analysis, even if listeners misperceive only one phone, it is considered as a word misperception as a whole. There might be possibility that listeners' partial word-word recognition accuracy could provide an evidence to support a neighborhood density effect. In order to examine this possibility, the phones correct analysis was first conducted.

Listeners' percent phonemes correct was submitted to a four-factor mixed-model ANOVA, with talker, frequency and density as the within-subjects factors, and group as the between-subjects factor. Significant effects of density ($F[1,66]=20.20$, $p<0.001$, $\eta^2_{\text{partial}}=0.23$), talker ($F[1,66]=178.01$, $p<0.001$, $\eta^2_{\text{partial}}=0.73$), and group ($F[2,66]=372.71$, $p<0.001$, $\eta^2_{\text{partial}}=0.99$) were observed. In addition, there were significant two-way interactions between frequency and talker ($F[1,66]=6.79$, $p=0.011$, $\eta^2_{\text{partial}}=0.03$), between talker and group ($F[2,66]=17.491$, $p<0.001$, $\eta^2_{\text{partial}}=0.35$), and between density and frequency ($F[1,66]=24.42$, $p<0.001$, $\eta^2_{\text{partial}}=0.27$). Further, there is also a three-way interaction between frequency, density and talker ($F[1,66]=19.47$, $p<0.001$, $\eta^2_{\text{partial}}=0.22$).

The similar patterns observed in the correct words analysis were also attested in this analysis. As shown in Fig. 3 and 4, for all participants, low density words were recognized more accurately than high density words. Words spoken by the English native speaker were recognized more accurately than the ones spoken by the Japanese native speaker. The Minneapolis English and the Minneapolis Japanese equally

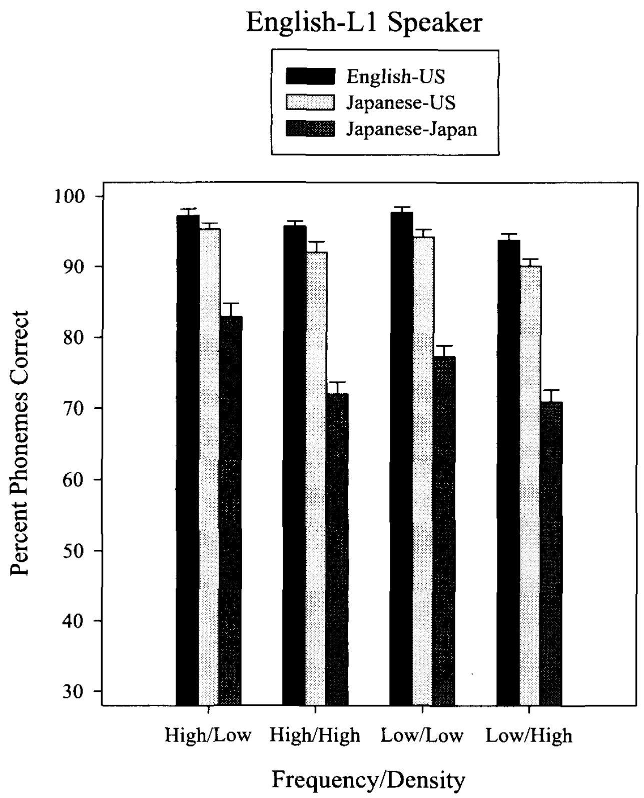

Fig. 3 Percent of phonemes in words produced by the native English speaker that were correctly identified, separated by listener group.

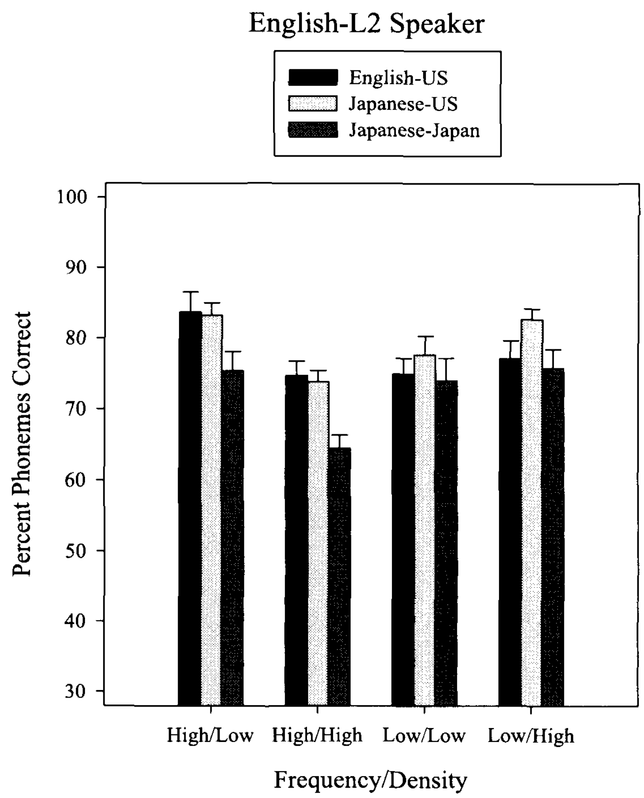

Fig. 4 Percent of phonemes in words produced by the native Japanese speaker that were correctly identified, separated by listener group.

recognized words accurately, and both recognized words more accurately than the Tokyo Japanese did. A large effect of frequency and a smaller effect of talker were observed for the Tokyo Japanese. A smaller effect of density was observed for low frequency words. The phones in the low frequency high density words were recognized less accurately when spoken by the native Japanese speaker than we would predict given the accuracy of phones in the low frequency high density words.

An interesting outcome was revealed by the phonemes correct analysis. The Japanese native speaker's productions of low-frequency, high-density English words were perceived more accurately than we would predict given the overall low intelligibility of this speaker. Since phoneme perception is also important in L2 word recognition, we decided to examine this finding further by examining the listeners' perceptual confusions for phonemes produced by the two talkers. This was done by tallying confusion matrixes for vowels and onset consonants, aggregated across the listeners, to determine which phones were accurately or inaccurately perceived by the participants. Here, we only consider the data of the Minnesota English listeners and the Tokyo Japanese listeners, since we have already learned that the Minnesota English and the Minnesota Japanese performed similarly. The listeners' written responses were converted to a phoneme equivalent using the Carnegie-Mellon pronouncing dictionary. As with the earlier analysis, vowel-plus-/r/ combinations were treated as a single diphthongal vowel. Complex onsets (i.e., clusters) were treated as a single unit and were compared both to other clusters and to singleton onsets.

First, let us look at confusion matrices for vowels. Tables 2 and 3 are vowel confusion matrices for the Minnesota English listeners hearing the English native speaker and for the Minnesota English listeners hearing the Japanese native speaker, respectively. As shown in Table 2, the Minnesota English listeners did not have any problem perceiving English vowels spoken by the English native speaker. In contrast, as shown in Table 3, for the words spoken by the Japanese native speaker, the Minnesota English listeners were not able to recover a postvocalic /r/ from the Japanese talker's productions of derhoticized /or/ and /ar/ in words like *corn* and *part*. This is not surprising since the Minnesota English listeners were from a rhotic dialect. Further, /ɔ/ was

Table 2 Vowel Confusion Matrix: Minneapolis English hearing English L1

	ɑ	æ	ʌ	aɪ	ɑr	aʊ	eɪ	ɛ	i	ɪ	o	ɔ	ɔɪ	or	u	ʊ	ɝ	other
ɑ	86	9	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0
æ	0	99	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
ʌ	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
aɪ	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ar	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0
aʊ	0	0	0	0	0	98	0	0	0	0	0	3	0	0	0	0	0	0
eɪ	0	0	0	0	0	0	98	0	0	0	0	0	0	0	0	0	0	2
ɛ	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0
i	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0
ɪ	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0
o	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0
ɔ	1	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0
ɔɪ	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0
or	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0
u	0	0	0	0	0	0	0	0	5	0	0	0	0	0	95	0	0	0
ʊ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	91	0	9
ɝ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	2

often confused with /ʌ/.

Tables 4 and 5 are vowel confusion matrices for the Tokyo Japanese listeners hearing the English native speaker and for the Tokyo Japanese listeners hearing the Japanese native speaker, respectively. Tables shows that for the Tokyo Japanese listeners, the vowels /ɑ/, /ɑr/, /ɔ/, /or/, and /ɝ/ were hardest to perceive in words recorded both by the English and Japanese native speakers. These vowels occurred in the following words, whose lexical characteristics are noted (HF=high frequency, LF=low frequency, HD=high density, LD=low density): /ɑ/: *job* (HFHD), *sock* (LFHD); /ɑr/: *park, part* (HFHD), *cart* (LFLD); /ɔ/: *call, fall*, (HFHD) *lost, wrong* (HFLD), *boss, frog, wash* (LFLD); /or/: *corn* (LFHD), *fork* (LFLD); /ɝ/: *heard* (HFHD), *bird, burn, hurt* (LFHD). Further, the vowel /o/ was hard to perceive in the words spoken by the Japanese native speaker. A further analysis revealed that 72% of the words with vowels that were overall hard to identify were from dense neighborhoods.

The most accurately recognized vowels for the Tokyo Japanese listeners were /aɪ/, /aʊ/, /i/, /eɪ/, /ɔɪ/, /u/ (/aɪ/: *five, kind, smile, white* (HFLD), *hide, shine* (LFHD); /aʊ/: *house, mouth* (HFLD), *loud, mouse* (LFLD); /i/: *peace, sheep* (LFHD), *cheese, teach* (LFLD); /eɪ/: *date, face, lake, rate, save* (HFHD), *faith, safe* (HFLD), *cake, nail, shake* (LFHD); /ɔɪ/: *choice, join, voice*, (HFLD) *coin, noise* (LFLD); /u/: *foot, move* (HFLD), *noon, soup* (LFHD)). 58% of the words with accurately identified vowels were from sparse neighborhoods (74% excluding /eɪ/).

A χ^2 test showed that well-recognized and poorly recognized vowels by the Tokyo Japanese were not distributed randomly among high- and low-density words ($\chi^2_{[df=1]}$=5.88, p=0.015). Therefore, it is difficult to determine whether the apparent effects of neighborhood density on Japanese listeners' responses were due to neighborhood density *per se*, or to the fact that the specific high- and low-density words used in this study happened to contain vowels that were intrinsically difficult or easy to perceive.

Consonant confusion matrices for the Minnesota English listeners hearing the English native speaker and for the Minnesota English listeners hearing the Japanese native speaker revealed that the Minnesota English listeners did not have any problem perceiving

Table 3 Vowel Confusion Matrix: Minneapolis English hearing English L2

	ɑ	æ	ʌ	aɪ	ɑr	aʊ	eɪ	ɛ	i	ɪ	o	ɔ	ɔɪ	or	u	ʊ	ɝ	other
ɑ	59	0	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
æ	1	82	12	1	3	1	0	0	0	0	0	0	0	0	0	0	0	0
ʌ	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
aɪ	0	0	0	98	0	0	0	0	0	0	0	0	0	0	0	0	0	2
ar	31	0	3	0	19	0	0	0	0	0	3	44	0	0	0	0	0	0
aʊ	0	0	2	2	2	88	0	2	0	0	0	4	0	0	0	0	0	0
eɪ	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
ɛ	0	12	0	0	0	0	3	62	0	12	0	0	0	0	1	0	0	10
i	0	0	0	0	0	0	0	0	84	16	0	0	0	0	0	0	0	0
ɪ	0	0	0	0	0	0	2	0	2	95	0	0	0	0	0	0	0	0
o	2	0	6	0	0	0	0	0	0	0	71	6	6	4	0	6	0	0
ɔ	6	3	19	0	0	0	0	0	0	0	15	47	1	4	0	3	0	1
ɔɪ	0	0	0	0	0	0	0	0	0	0	0	0	98	0	2	0	0	0
or	13	0	0	0	0	0	0	0	0	0	46	17	4	8	0	13	0	0
u	0	0	0	0	0	0	0	0	0	0	0	0	0	0	95	0	2	2
ʊ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
ɝ	5	0	0	0	26	0	0	0	0	0	0	2	0	0	0	2	62	2

consonants in the words spoken by the English native speaker since /br/ was the only one which did not reach to the 80% accuracy level, an effect that was driven solely by some listeners misperceiving the word *bring* as *green.* On the contrary, 8 out of 21 onset consonants were perceived below the 80% accuracy level for the words spoken by the Japanese native speaker. Consonant confusion matrices for the Tokyo Japanese listeners hearing the English native speaker and for the Tokyo Japanese listeners hearing the Japanese native speaker also revealed that the Tokyo Japanese listeners consistently presented their difficulty perceiving /d/, /f/, /fr/, /h/, /p/, /r/, /v/, /w/ in the words spoken by the English and Japanese native speakers. This may indicate the sound perception difficulty specifically for the second language learners of English. In contrast to vowels, there was no systematic relationship between consonant errors and the lexical characteristics of the stimuli. The consonant errors appeared to follow well-known differences between the languages' phoneme inventories.

3.3 Relating Familiarity to Accuracy

The next analysis examined group differences in rated familiarity of the words in the experiment, as well as the relationship between familiarity and word-recognition accuracy. First, we examined the influence of word frequency and phonological neighborhood density on familiarity ratings for the three groups. Listeners' ratings were submitted to a three-factor mixed-model ANOVA, with frequency and density as the within-subjects factors, and group as the between-subjects factor. Significant effects of frequency ($F[1,65]=91.4$, $p<0.001$, $\eta^2_{partial}=0.58$), density ($F[1,65]=8.7$, $p=0.004$, $\eta^2_{partial}=0.12$), and group ($F[2,65]=7$, $p=0.002$, $\eta^2_{partial}=0.18$) were observed. In addition, there was a significant interaction between frequency and neighborhood density ($F[1,65]=8.5$, $p=0.008$, $\eta^2_{partial}=0.10$). As shown in Fig. 5, low-frequency words were rated as less familiar than the high-frequency words. Moreover, high-density words were rated as less familiar than low-density ones, though this tendency was only present for the low-frequency words. Surprisingly, there was an inverse relationship between English proficiency and perceived familiarity: the least proficient speakers, the Tokyo Japanese rated the items as most familiar.

Table 4 Vowel Confusion Matrix: Tokyo Japanese hearing English L1

	ɑ	æ	ʌ	aɪ	ɑr	aʊ	eɪ	ɛ	i	ɪ	o	ɔ	ɔɪ	or	u	ʊ	ɝ	other
ɑ	48	39	0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0
æ	0	76	6	0	0	1	1	6	1	0	0	0	0	0	0	0	6	1
ʌ	17	9	69	0	0	3	0	0	0	0	0	3	0	0	0	0	0	0
aɪ	0	0	0	87	0	0	10	0	0	3	0	0	0	0	0	0	0	0
ar	3	6	18	0	55	0	0	0	0	0	0	12	0	6	0	0	0	0
aʊ	0	0	0	0	0	89	0	0	0	0	5	2	0	5	0	0	0	0
eɪ	0	3	0	0	0	0	87	0	7	3	0	0	0	0	0	0	0	0
ɛ	0	11	0	0	0	1	1	76	0	0	10	0	0	0	0	0	1	0
i	0	0	0	0	0	0	2	0	87	11	0	0	0	0	0	0	0	0
ɪ	0	0	0	0	0	0	0	26	0	70	2	0	0	0	0	2	0	0
o	2	0	2	0	0	10	0	0	0	0	78	5	0	2	2	0	0	0
ɔ	9	4	9	0	5	0	0	0	0	0	14	51	0	3	3	3	1	0
ɔɪ	0	0	0	0	0	0	0	2	0	0	0	0	98	0	0	0	0	0
or	18	0	0	0	0	0	0	0	0	0	5	9	0	45	0	23	0	0
u	0	0	4	0	0	0	0	0	0	11	0	0	0	0	85	0	0	0
ʊ	13	0	0	0	0	0	0	0	0	0	0	4	0	0	4	78	0	0
ɝ	2	9	0	0	19	0	0	0	0	0	0	0	0	2	2	9	57	0

We also examined the difference between average familiarity ratings for real words (pooled across the four word types) and the nonword foils for the three groups in the two-factor mixed-model ANOVA with lexicality (real word vs. nonword) as the within-subjects factor and subject group (Minnesota English, Minnesota Japanese, Tokyo Japanese) as the between-groups factor. Significant main effects were found for lexicality (F[1,65]=2265.7, p<0.001, $\eta^2_{partial}$=0.97), group (F[2,65]=9.6, p<0.001, $\eta^2_{partial}$=0.23), and an interaction between them (F[2,65]=3.83, p=0.027, $\eta^2_{partial}$=0.11). As shown in Fig. 6, again, there was an inverse relationship between English proficiency and perceived familiarity for both real words and non-word foils: the least proficient speakers, the Tokyo Japanese rated the items as most familiar.

The last analysis examined the association between familiarity and both word- and phoneme-recognition accuracy. Spearman's rho correlations were calculated between rated familiarity and both word- and phoneme-recognition accuracy for all 68 listeners, as one of the Tokyo Japanese listeners did not complete the familiarity task. Of these, 50% (11/22) of the Tokyo Japanese listeners, 33% (8/24) of the Minnesota Japanese listeners, and 5% (1/22) of the Minnesota English listeners had significant correlations between at least one of the recognition accuracy scores and rated familiarity. However, inspection of these individual correlations showed that the listeners with these correlations were also those who were less accurate overall.

4. Summary and Discussion

This study aimed to answer two questions. The first question tried to reveal whether neighborhood density, word frequency, and/or word familiarity affect word recognition in English with Japanese L1 speakers' acquisition of English as an L2. The second question invested whether apparent effects of phonological neighborhood density can be attributed to the phonetic content of the stimuli rather than to lexical characteristics. In order to answer these questions, a similar experiment in Imai et al. (2005) was conducted. The results of this experiment provided following findings.

Related to the first question, we found that indeed there were strong effects of frequency and neighborhood

Table 5 Vowel Confusion Matrix: Tokyo Japanese hearing English L2

	ɑ	æ	ʌ	aɪ	ɑr	aʊ	eɪ	ɛ	i	ɪ	o	ɔ	ɔɪ	or	u	ʊ	ɝ	other
ɑ	57	9	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
æ	4	72	12	0	0	1	0	0	0	0	2	1	0	0	0	0	7	0
ʌ	0	6	94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
aɪ	0	0	0	88	0	0	4	1	0	4	0	0	1	0	0	0	0	0
ar	22	22	17	0	6	3	3	0	0	0	8	14	0	3	0	0	3	0
aʊ	0	0	2	0	2	90	0	0	0	0	0	4	0	2	0	0	0	0
eɪ	0	0	0	0	0	0	95	0	4	0	1	0	0	0	0	0	0	0
ɛ	0	7	0	0	0	0	2	70	1	5	2	0	0	0	5	1	1	4
i	0	0	0	0	0	0	0	0	87	13	0	0	0	0	0	0	0	0
ɪ	0	0	0	0	0	0	0	2	4	93	0	0	0	0	0	0	0	0
o	5	0	0	0	0	7	0	0	0	0	63	14	2	7	2	0	0	0
ɔ	9	17	7	0	0	1	0	0	0	0	19	43	0	1	1	1	0	0
ɔɪ	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0
or	4	0	4	0	0	0	0	0	0	0	25	17	4	42	4	0	0	0
u	0	0	4	0	0	2	0	0	0	0	4	0	0	0	83	4	2	0
ʊ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
ɝ	7	4	4	0	20	0	0	0	0	0	9	9	0	0	2	9	33	2

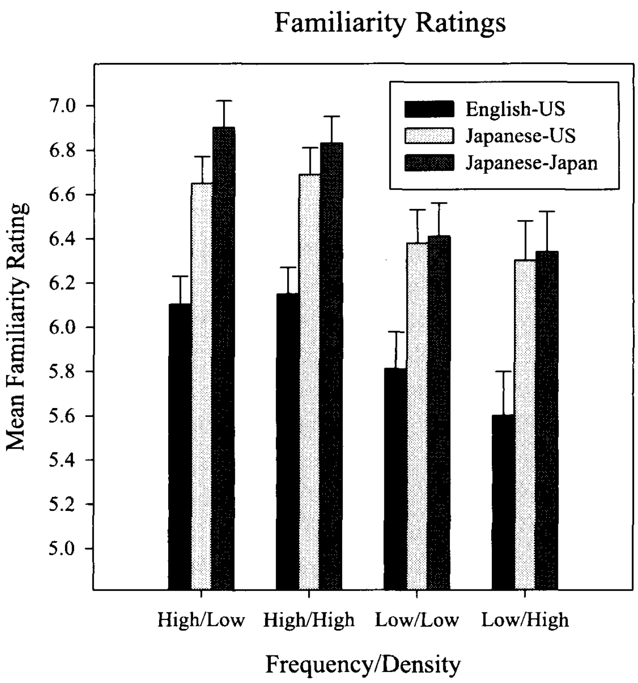

Fig. 5 Average rated familiarity for the real words separated by word frequency and phonological neighborhood density for the three groups of listeners.

density on the performance of all three groups of listeners: beginning Japanese L2 English speakers living in Tokyo, advanced Japanese L2 English speakers living in Minnesota, and native English speakers living in Minnesota. However, there was no clear evidence for an emerging 'neighborhood competition' effect in the Japanese learners of English. That is, we didn't see a reduced competition effect in the Tokyo Japanese listeners relative to the Minneapolis Japanese listeners and the Minneapolis English listeners. Unlike in Imai et al., who did show a bigger discrepancy between high- and low-density words in more- and less-proficient learners of English whose L1 was Spanish.

However, the results suggest there was at least one confound in the stimuli that may have mediated these results. As discussed in the methods section, the distribution of phonemes was not random across different frequencies and densities. The distribution of onset consonants differs marginally between high- and low-density words. Vowel nuclei differed significantly between high- and low-density words. The distribution of codas differed significantly as a function of both density and, marginally word frequency. Moreover, unlike

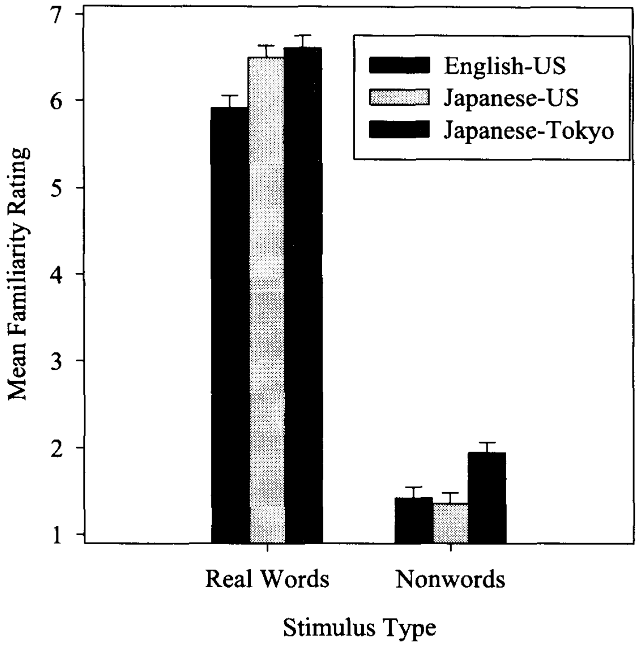

Fig. 6 Average rated familiarity for real words (averaged across all 80 words) and nonword foils, separated by listener group.

Imai et al., we separated our groups based on location, not on proficiency. This was done for expedience—we are currently measuring the proficiency and we re-analyze the data when we have these findings. Even so, there is a high probability that effects of phonological neighborhood density in this study were not attributed to lexical characteristics.

Related to the second question, the phoneme accuracy results suggest that apparent effects of phonological neighborhood density can potentially be attributed to the influence of the phonemic composition of the stimuli on listeners' perceptions. Difficult sounds for the listeners occurred more frequency in the words from dense neighborhood than in the ones from sparse neighborhoods, whereas easy sounds that were easy for the listeners to perceive occurred more frequently in the words from sparse neighborhoods than in the ones from dense neighborhoods. The results clearly indicate that L2 word recognition is highly dependent on listeners' phoneme perception, which could lead to apparent effects of the lexical properties for this stimulus set.

The results of this study suggest that we need a new methodology for examining the relative contribution of word- and phoneme-level representations on L2 performance. One possibility that we are currently examining was inspired by the literature on word recognition in individuals with hearing impairment. Studies of the factors that affect word recognition in that population have a problem that closely parallels that which we faced: inaccurate of recognition of a word like *soothe* can't be unambiguously attributed to something like low frequency of usage, as the component sounds in this word—particularly the fricatives /s/ and /ð/—are often very difficult for listeners with even a mild hearing impairment to perceive. One solution to this is presented by Boothroyd and Nittrouer (1988, see also Nittrouer and Boothroyd 1990 and Benkí 2003). If independent measures of word recognition accuracy and recognition of the sounds that comprise these words are taken, then log-linear models can be used to examine the discrepancy between actual word-recognition accuracy and the accuracy that would be predicted by listeners' accuracy in perceiving the component phones. Using this method, Nittrouer and Boothroyd (1990) showed that young children's measured word recognition accuracy was very close to what would be predicted from their perception of the phonemes comprising the words, while older adults' word recognition was better than would be predicted by phoneme recognition. This suggests that children do not robust enough knowledge of the phonemic structure of words to recognize words based on incomplete information. Benkí (2003) showed that phonological neighborhood density and frequency affected adult listeners' recognition of spoken words even when phoneme-perception accuracy was controlled statistically. Applying Benkí's method to the study of L2 acquisition will provide a clearer picture of the relationship among L2 proficiency, lexical factors, and spoken-word recognition than the current methods allow.

Acknowledgments

This paper is a revised version of the paper that was presented at the International Phonetics and Phonology Forum held in Kobe, Japan between August 26 and 27, 2009. We thank Sho Haraguchi and other organizing committee members who invited us for a talk at the Forum. We further thank Takashi Otake, chair of our session at the Forum, for valuable comments on the earlier version of the paper. Our appreciation also goes to Kari Urberg-Carlson and Celina Marnie, for their testing in Twin Cities, and people who participated in the study in Twin Cities and Tokyo. This study was partly supported by JSPS Grand-in-Aid for Young Scientists (B) 18720110 to Kiyoko Yoneyama, and by a grant from the US National Science Foundation (BCS-0729277) to Benjamin Munson.

References

Benkí, J.R. (2003) "Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition," *Journal of the Acoustical Society of America* 113, 1689–1705.

Best, C.T. (1995) "A direct realist view of cross-language speech perception." In W. Strange (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* (pp.171–206). Baltimore: York Press.

Boersma, P. (2001) "Praat, a system for doing phonetics by computer," *Glot International* 5, 341–345.

Boothroyd, A and Nittrouer, S. (1988) "Mathematical treatment of context effects in phoneme and word recognition," *Journal of the Acoustical Society of America* 84, 101–114.

Bradlow, A.R. and Pisoni, D.B. (1999) "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," *Journal of the Acoustical Society of America* 112, 272–284.

Charles-Luce, J. and Luce, P.A. (1990) "An examination of similarity neighborhoods in young children's receptive vocabularies," *Journal of Child Language* 22, 727–735.

Flege, J. and Liu, S. (2001) "The effect of experience on adults' acquisition of a second language," *Studies in Second Language Acquisition* 23, 527–552.

Flege, J.E., Munro, M.J. and MacKay, I.R.A. (1995) "Factors affecting degree of perceived foreign accent in a second language," *Journal of Acoustical Society of America* 97, 3125–3134.

Gahl, S. (2008) ""Thyme" and "Time" are not homophones. Word durations in spontaneous speech," *Language* 84:3, 474–496.

Imai, S., Walley, A.C. and Flege, J.E. (2005) "Lexical frequency and neighborhood density effect on the recognition of native and Spanish-accented words by native English and Spanish listeners," *Journal of the Acoustical Society of America* 117, 896–907.

Li, F. (2008) *The phonetic development of voiceless sibilant fricatives in English, Japanese, and Mandarin Chinese.* Doctoral dissertation, Department of Linguistics, Ohio State University.

Luce, P.A. and Pisoni, D.B. (1998) "Recognizing spoken words: The neighborhood activation model," *Ear and Hearing* 19, 1–36.

MacKay, I.R.A., Meador D. and Flege, J.E. (2001) "The identification of English consonants by native speakers of Italian," *Phonetica* 58, 103–125.

Meador, D., Flege, J.E. and MacKay, I.R.A. (2000) "Factors affecting the recognition of words in a second language," *Bilingualism: Language and Cognition* 3, 55–67.

Metsala, J.L. and Walley, A.C. (1998) "Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability." In J.L. Metsala and L.C. Ehri (eds.), *Word recognition in beginning literacy.* (pp.89–120). NewYork: Erlbaum.

Munson, B. and Solomon, N.P. (2004) "The effect of phonological neighborhood density on vowel articulation," *Journal of Speech, Language, and Hearing Research* 47, 1048–1058.

Munson, B., Swenson, C.L. and Manthei, S. (2005) "Lexical and phonological organization in children: evidence from repetition tasks," *Journal of Speech, Language, and Hearing Research* 48, 108–123.

Nittrouer, S. and Boothroyd, A. (1990) "Context effects in phoneme and word recognition by young children and older adults," *Journal of the Acoustical Society of America* 87, 2705–2715.

Sebastian-Galles, N. and Soto-Faraco, S. (1999) "Online processing of native and non-native phonemic contrasts in early bilinguals," *Cognition* 72, 111–123.

Sekiguchi, T. (2006) "Effects of Lexical Prosody and Word Familiarity on Lexical Access of Spoken Japanese Words," *Journal of Psycholinguistic Research* 35, 369–384.

Storkel, H.L. (2002) "Restructuring similarity neighborhoods in the developing mental lexicon," *Journal of Child Language* 29, 251–274.

Vitevitch, M.S. and Luce, P.A. (1999) "Probabilistic phonotactics and neighborhood activation in spoken-word recognition," *Journal of Memory and Language* 40, 374–408.

Vitevitch, M.S. and Rodríguez, E. (2005) "Neighborhood density effects in spoken word recognition in Spanish," *Journal of Multilingual Communication Disorders* 3, 64–73.

Vitevitch, M.S. and Stamer, M.K. (2006) "The curious case of competition in Spanish speech production," *Language & Cognitive Processes* 21, 760–770.

Yamada, R. (1995) "Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese." In W. Strange (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* (pp.305–320). Baltimore: York Press.

Yao, Y. (2010) "Separating talker- and listener-oriented forces in speech using phonological neighborhood density," Oral presentation at the Annual Meeting of the Linguistic Society of America, Baltimore, January 7.

Yoneyama, K. (2002) *Phonological neighborhoods and phonetic similarity in Japanese word recognition.* Doctoral dissertation, Department of Linguistics, Ohio State University.

(Received Feb. 5, 2010, Accepted May 12, 2010)